

Occam's Razor Needs New Blades*

by

Herman Rubin

Department of Statistics

Division of Mathematical Sciences

Mimeograph Series No. 216

December 1969

*This research was partly supported by the Office of Naval Research Contract N00014-67-A-0226-0008. Reproduction in whole or in part is permitted for any purpose of the United States Government.

Occam's Razor Needs New Blades*

by

Herman Rubin

The principle was announced by William of Occam in the middle ages to be used in theology that "one should not multiply causes without reason." This principle has been adopted by natural philosophers and made a fundamental principle of scientific inference. It is not clear exactly what this statement means in scientific problems. However, certain procedures have been adopted on a possible interpretation of this statement, and in this paper we intend to examine these procedures and to show that they are not valid applications of Occam's Razor. There still remains a problem as to exactly what Occam's Razor means for scientific purposes. We hope to throw some light on this problem and that our observations will lead to a more accurate formulation of the problem of scientific inference.

A blade which has been extensively used in scientific pursuits is to assign a significance level and to test a null hypotheses, usually against a parametric alternative. In some cases the parametric alternative is nothing more than a change in the number of powers of a variable or variables which it has already been decided to include as causes. This latter is not a matter of consideration of exclusion of causes, but exclusion of the complexity of causes, which seems somewhat related, though different from the original principle. However, even when the matter of

*This research was partly supported by the Office of Naval Research Contract N00014-67-A-0226-0008. Reproduction in whole or in part is permitted for any purpose of the United States Government.

consideration of which variables to include is encountered, one normally considers including these variables because one has good reason already to believe those variables are causes in the scientific problem under consideration. There are a few problems, like that of the existence of extra-sensory perception or the constancy, of the velocity of light in vacuum, where one seriously considers the null hypotheses. However, in other situations, like whether teaching machines have an effect on the performance of students, or whether cloud seeding has an effect on the total amount of rainfall, where the problem is not so much the existence of the cause, but of the size of the effect, and whether there is any practical importance in including the cause.

However, let us now consider the case when one is really interested in testing whether or not the cause occurs. Two things should be kept in mind. First, we are frequently deciding not whether there is some cause to be included, but whether a particular cause, which we have some reason to believe should be included. Second, we should keep in mind that the test whether the cause should be included is affected by the correctness of our theory. Nevertheless, in this case, if the sample size is fixed and the sample is not too small, the standard statistical tests are appropriate tests to use. In using these tests there is the problem of deciding what the significance level should be. It has been observed by many authors that the significance level should change with sample size and, in fact, should generally decrease as the sample size increases. As a corollary of the decrease with increasing sample size there should be, of course, an increase with decreasing sample size, so that for very small sample sizes it may be that one should not even consider accepting the null hypotheses no matter what the data is! Several authors have

obtained methods of evaluating approximate significance levels, based on the user's assessment of risks. In no case can the appropriate significance level be determined in an "objective" manner.

The main problem in scientific inference is that of deciding when to "accept" or to "announce" a theory. By acceptance of a theory, I mean the taking of a position that, at the present time, it is desirable to proceed as if the theory were true. In many branches of astronomy, the Newtonian theory of gravitation is accepted. There is even a secondary type of acceptance, namely the taking of a position that the action of the "main causes" is described by the theory, and that is desirable to try to further understand the theory. An example of this is Boyle's Law, or Kepler's Laws, which are somewhat crude approximations to the presently accepted theories, but the study of which leads to much of the present development.

The announcement of a theory is the taking of a position that, at the present time, it is desirable to proceed as if the theory might be true. This is the situation, for example, as regards the various approaches to general relativity and cosmology, and in a great many situations in the behavioral sciences.

In both of these situations, the action to be taken cannot even be in principle forced by the data only. A theory which, on certain data, is accepted today, may, on the same data, be considered tomorrow as merely an approximation.

One may ask why it is necessary to accept or announce theories in which one could not believe. This is so as to enable the making of predictions, and the attainment of understanding. Without any theory, nothing could be done -- even the cataloging of data requires the acceptance

of a theory. The simpler a theory, the more likely it is to lead to understanding.

For this use of Occam's Razor, the appropriate statistical blades (which are necessary except in the simple situations which have prevailed in the physical sciences and occasionally in the biological sciences) have not yet been forged. I believe the forging of these blades will involve the cooperation of theoretical scientists and theoretical statisticians who are as far as possible unprejudiced by their exposure to classical procedures.

There is yet another case in which Occam's Razor is mistakenly used; however, here an appropriate carving knife is available. This is the problem of one-sided testing. Here for moderately large samples, it is only necessary to ask whether the expected gain of introducing the new procedure will outweigh its expected costs, including the necessarily a priori assessment of as yet unobserved side effects. The assessment of the prior distribution of nature is relatively unimportant, as is typically the case in estimation.

In summary, the use of α - level significance tests with fixed α as a tool for inference does not seem to have any justification as an application of Occam's Razor. The author has been unable to find any validity for this use except in the certain one-sided cases where, the appropriate significance level is usually approximately one-half.